CMIP Status and Infrastructure Requirements

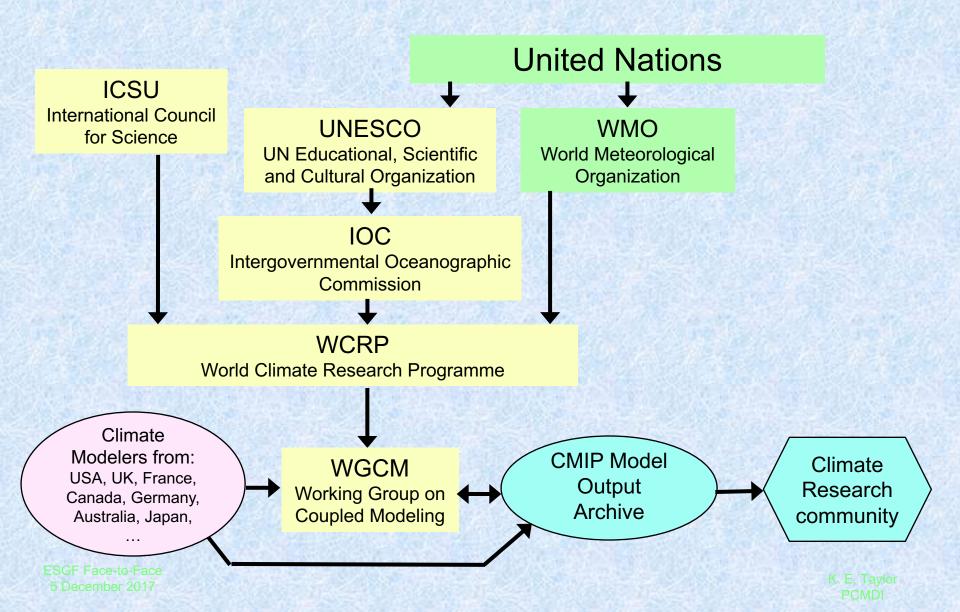
Karl E. Taylor and V. Balaji

on behalf of the WGCM Infrastructure Panel

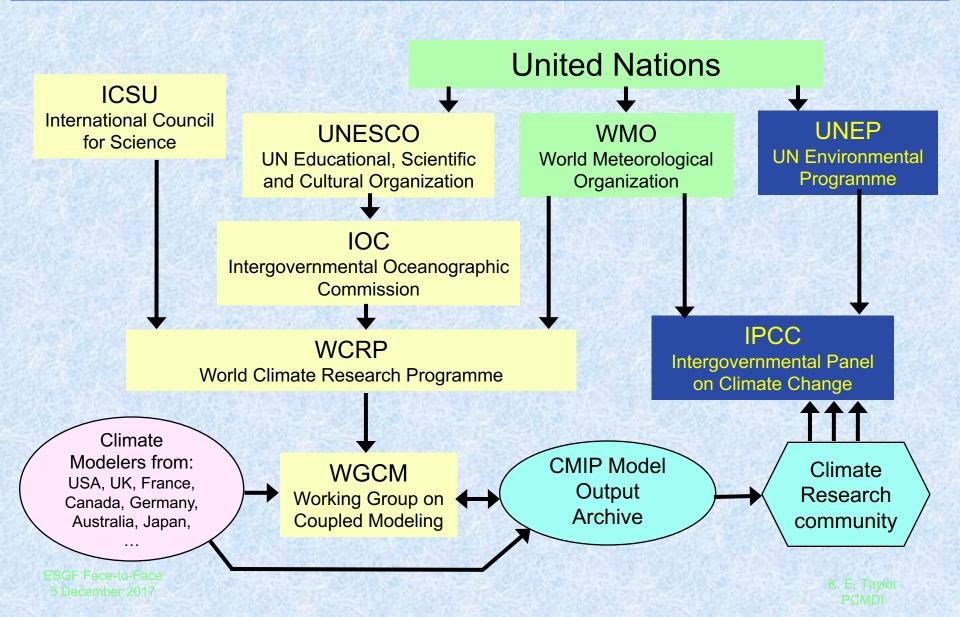
Presented at the 2017 Earth System Grid Federation Face-to-Face Conference

San Francisco, CA 5 December 2017

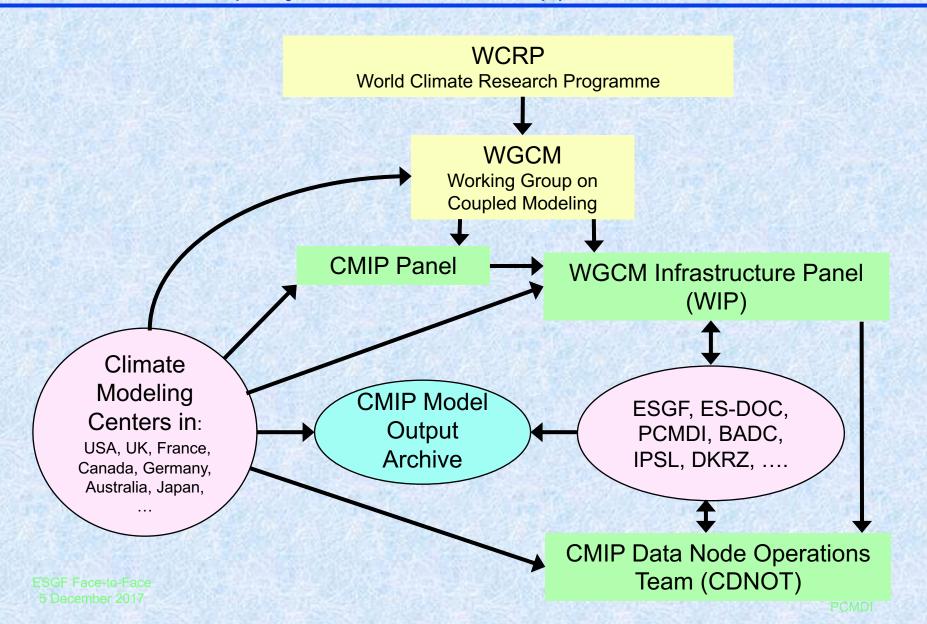
CMIP's place in international climate science



IPCC assessments are separate from the international climate research programs



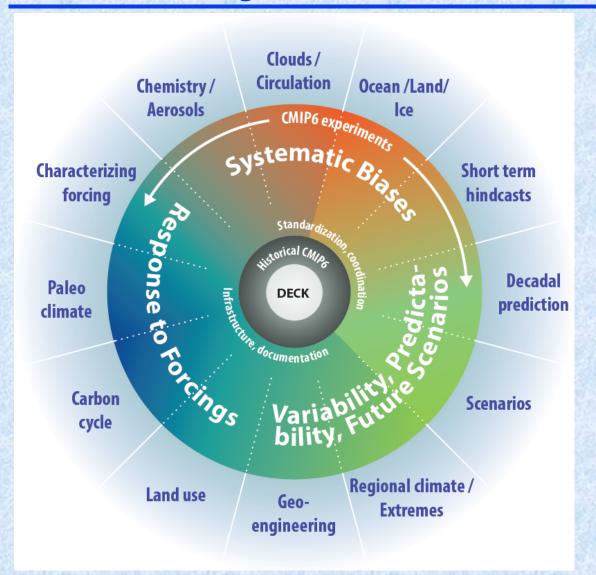
The WGCM appoints panels to coordinate its work and the funded projects makes it happen.



CMIP infrastructure coordination

- The WGCM Infrastructure Panel (WIP) appointed by the WGCM
- The WIP
 - Manages and coordinates infrastructure development and implementation.
 - Maintains a website hosting "Position Papers" and some of the specifications for CMIP6:
 - https://www.earthsystemcog.org/projects/wip/
 - Oversees the CMIP Data Node Operations Team (CDNOT)
- Major contributions to the infrastructure come from ESGF, ES-DOC, PCMDI, BADC, IPSL, DKRZ, and others.

CMIP6 design overview:



DECK

- Small set of benchmark runs
- To evolve only slowly (e.g. OMIP, LMIP)

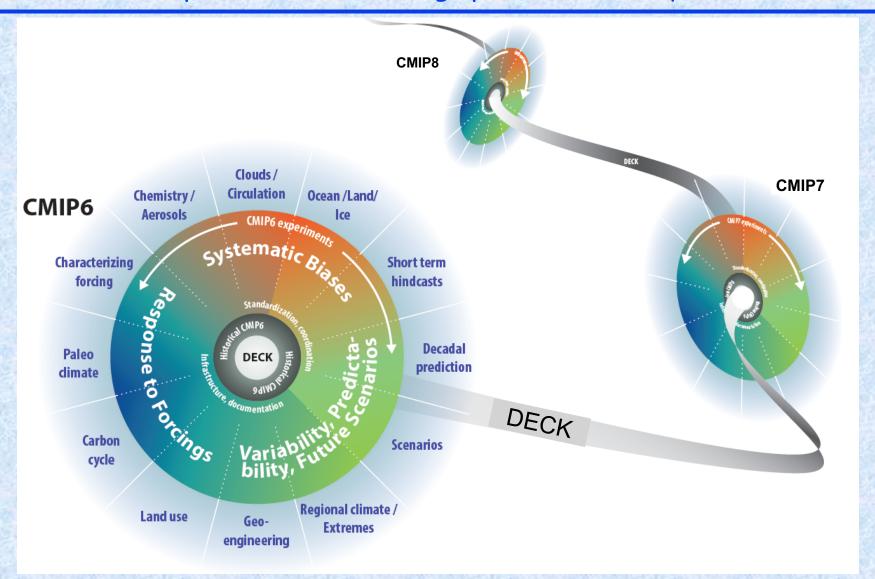
Historical CMIPX

 Forcing to be updated for each new phase

CMIP6-endorsed MIPs

 An evolving collection to address specific scientific issues

CMIP provides continuity through DECK and an evolving suite of additional experiments addressing specific science questions.



More institutions, more models, more experiments, more data

- 33 institutions/consortia have officially registered for CMIP6
- 75 models/source_id's are registered
- 248 experiments
- order 20 PB of model output expected

CMIP6_CVs

https://github.com/WCRP-CMIP/CMIP6_CVs

Core Controlled Vocabularies (CVs) for use in CMIP6

Registering Institutions, Models, or requesting changes to CVs:

To register your institution or model or to request changes to a CV, please submit an issue/ticket following the instructions on the CMIP6_CVs issue page.

Some support for CMIP participating modeling groups is available: pcmdi-cmip@llnl.gov

To view the current experiment_id entries point your browser to CMIP6_experiment_id.html

To view the current institution_id entries point your browser to CMIP6_institution_id.html

To view the current source_id entries point your browser to CMIP6_source_id.html

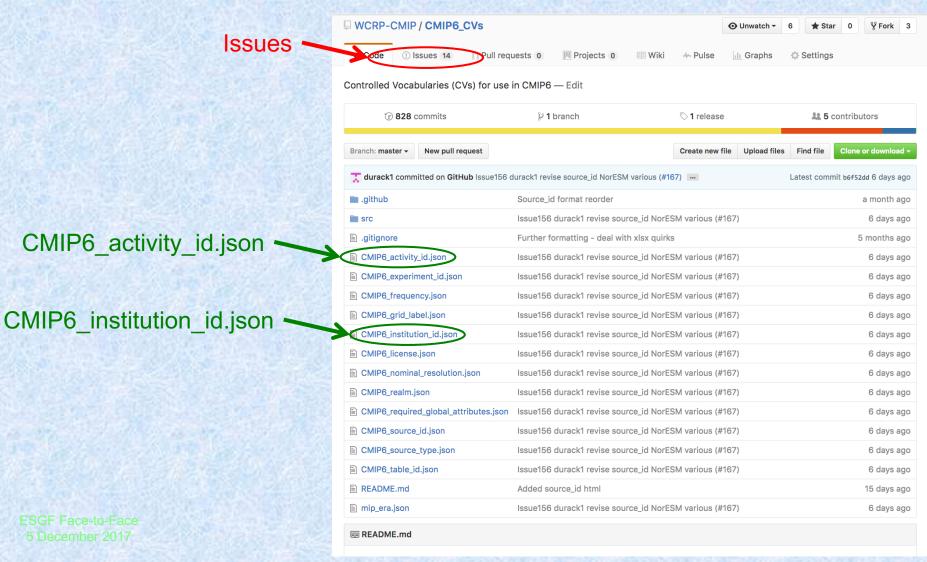
The CVs build on logic that is described in the CMIP6 Global Attributes, DRS, Filenames, Directory Structure, and

December CV's document

l aylo

Controlled vocabularies are specified in JSON files hosted by github

https://github.com/WCRP-CMIP/CMIP6_CVs



CMIP6 timeline

- A few groups have begun processing model output
- Modeling groups will try to complete most of their CMIP6 simulations during the next
 2 years
- CMIP6 is not tied to the IPCC AR6 timeline, but modeling groups are aware of the IPCC deadlines

THE PROPERTY OF THE PARTY OF TH		SEALS.	
2018			
4 Feb	Decision on selection of Authors		
25 June-1 July	First Lead Author Meeting (LAM1)		
2019			
7-13 Jan	Second Lead Author Meeting (LAM2)		
29 May-16 June	First Order Draft (FOD) Expert Review	Mary 1	
26 Aug-1 Sept	Third Lead Author Meeting (LAM3)		
2020			31 January 2020:
9 March-19 April	Second Order Draft Government and Expert Review		Journal articles submitted
1-7 June	Fourth Lead Authors Meeting (LAM4)		15 October 2020:
7 Dec - 31 Jan	FGD Government Review of Summary for Policy Makers		Journal articles accepted
2021			
12-18 April	IPCC Acceptance / Adoption / Approval		K. E. Taylor

What modeling groups need?

- Description of each experiment
 - Geoscientific Model Development special issue (Veronika Eyring, Ed.)
 - https://www.geosci-model-dev.net/special_issue590.html
- List of output fields requested from each experiment
- Forcing data sets
- Software to help meet CMIP6 data standards and check conformance

CMIP data request tools and documentation (Martin Juckes)

Information available at the WIP CoG site:

https://www.earthsystemcog.org/projects/wip/CMIP6DataRequest

CMIP6 Data Request

The CMIP6 experimental design and organization has been agreed at the WGCM 18th Session in October 2014, see details on the CMIP Panel website at http://www.wcrp-climate.org/index.php/wgcm-cmip/about-cmip. Part of this covers the creation and timeline of the CMIP6 Data Request.

The data request is available through a repository, and the latest version is available here (updated October 21st, 2016):

http://proj.badc.rl.ac.uk/svn/exarch/CMIP6dreq/tags/latest

An overview of the pressure levels proposed for atmospheric diagnostics is available for discussion (here).

Key documents describing the request (in the "docs" directory of the repository) are:

- Examples
- Python Library (dregPy)
- The Request XML document and Schema
- · Spreadsheet view of the variable definitions
- · A searchable list of variables in the request, linking to
- A browsable HTML view of the request
- Overview tables for tier 1, priority 1 and all tiers and priorities
- Discussion of issues: old forum, new github pages
- Registration for email list: CMIP6-DATAREQUEST@JISCMAIL.AC.UK
- Installation and usage of the python package

When problems are found, raise an issue! "CMIP6_DataRequest_VariableDefinitions"

See Version 01.beta.38 Release Notes for more details

What modeling groups need?

- Description of each experiment
- List of output fields requested from each experiment
- Clearly defined specifications of model output
 - Most important of metadata will be stored as global attributes
 - Controlled vocabularies will make it possible to interpret metadata
- Forcing data sets
- Software to help meet CMIP6 data standards and check conformance

Further information about data requirements:

- Reference "controlled vocabularies" (CV's) for CMIP6
 - https://github.com/WCRP-CMIP/CMIP6_CVs
- <u>Specifications</u> for file names, directory structures, and CMIP6
 Data Reference Syntax (DRS)
 - http://goo.gl/v1drZl
- Specification of output file content, structure, and metadata
 - not yet available,
 - with notable exceptions will follow CMIP5 requirements.
 - Use of CMOR3 will ensure compliance

46 Global attributes are defined in a table (with notes)

The attributes provide critical information needed to interpret the model output and are key attributes are relied on by the infrastructure.

CMIP6 global attribute see note 1	description	examples	corresponding attribute in CMIP5	form see note 2	when required?	further information and rationale
activity_id	activity identifier(s)	"CMIP", "PMIP", "LS3MIP LUMIP" see note 3	project_id	CV	always	renamed more generically, since not all activities are projects; also multiple activities may now be listed separated by single spaces.
branch_method	branching procedure	"standard", "none provided", "no parent" see note 4	-	free form	whenever parent exists	in CMIP6 some branching methods will involve short spin-up periods or other non-standard procedures which need to be described. See note 4. If no parent, omit or set to "no parent"
branch_time_in_ child	branch time with respect to child's time axis	365.0D0, 0.0D0 see note 5	-	double precision float	whenever parent exists	aids in interpreting branch times; units are the same as the units used for the child's time axis. If no parent, omit (preferred) or set to start time of the run.
branch_time_in_ parent	branch time with respect to parent time axis	3650.0D0 see note 5	branch_time	double precision float	whenever parent exists	changed name to explicitly distinguish it from branch_time_in_child; units are specified in the attribute: parent_time_units. If no parent, omit or set to 0.0D0.

What modeling groups need?

- Description of each experiment
- List of output fields requested from each experiment
- Clearly defined specifications of model output
- Forcing data sets
 - Input4MIPs
- Software to help meet CMIP6 data standards and check conformance

Forcing datasets for CMIP6: Input4MIPs status

- Project initiated April 2016
- Purpose
 - To collect, version-control, and archive CMIP6 forcing data sets
 - To impose data and metadata standards facilitating use
- Forcing datasets description/status
 - https://esgf-node.llnl.gov/projects/input4mips/
- input4MIPs holdings to-date
 - 1647 files & 559 Gb of data
 - 3 CMIP panel releases: v6.0.0, v6.1.1 and v6.2.0
 - Data footprint expected to ~double+ over coming months (satellite MIP data)
- input4MIPs project has adopted the CMIP infrastructure

Input4MIPs DECK/historical forcing data status

Forcing Dataset	Status	Temporal Coverage	Latest Data Version(s)	Contact	
SLCF Emissions	Available	1750-01 to 2014-12	2017-05-18, 2017-08-30 (Aircraft; updated)	Steven Smith <u>ssmith@pnnl.gov</u>	
Biomass Burning	Available	1750-01 to 2015-12	1.2 (2016-12-13; updated)	Margreet van Marle m.j.e.van.marle@vu.nl	
CO2 and CH4 Emissions	Available	1750-01 to 2014-12	2017-05-18, 2017-08-30 (Aircraft; updated)	Steven Smith <u>ssmith@pnnl.gov</u>	
Land-use	Available	850 to 2015	2.1h (2017-01-26)	George Hurtt <u>gchurtt@umd.edu</u>	
GHG concentrations	Available	0-01 to 2015-12	1.2.0 (2016-07-01)	Malte Meinshausen malte.meinshausen@unimelb.edu.au	
Ozone concentrations	Available	1850-01 to 2014-12	1.0 (2016-07-11)	Michaela Hegglin m.i.hegglin@reading.ac.uk	
Nitrogen deposition	Available	1850-01 to 2014-12	2.0 (2016-12-07; updated)	Michaela Hegglin m.i.hegglin@reading.ac.uk	
Simple plume aerosol	Available	1850 to 2100	1.0 (2017-02-01)	Bjorn Stevens bjorn.stevens@mpimet.mpg.de	
Solar	Available	1850-01 to 2299-12	3.2 (2017-01-03; updated)	Katja Matthes kmatthes@geomar.de	
Stratospheric aerosol	Available	1850-01 to 2014-12	3.0 (2017-10-04; updated)	Beiping Luo beiping.luo@env.ethz.ch	
AMIP SST and SIC	Available	1870-01 to 2016-06	1.1.2 (2017-04-19; updated)	PCMDI pcmdi-cmip@llnl.gov	

Download links, input4MIPs website: https://esgf-node.llnl.gov/search/input4mips
Also see the live google doc at https://goo.gl/r8up31

Endorsed-MIP forcing status

Satellite MIP	Status	Host(s); Version Committed to input4MIPs		Contact	
CFMIP	Ready	http://doi.org/10.5194/gmd-2016-70; 1.0	Ş	Mark Webb mark.webb@metoffice.gov.uk	
DAMIP	Ready	1.0 (2017-08-14)	-	David Plummer <u>david.plummer@canada.ca</u>	
DCPP	Ready	1.1 (2017-01-23)	-	Christophe Cassou christophe.cassou@cerfacs.fr	
FAFMIP	Ready	http://www.met.reading.ac.uk/~jonathan/FAFMIP/; (2015-08-21)	Yes	Jonathan Gregory j.m.gregory@reading.ac.uk	
HighResMIP	Ready	2.2.0.0-r1 (2017-05-05)	-	Malcolm Roberts malcolm.roberts@metoffice.gov.uk	
LS3MIP	Unknown			Sonia Seneviratne sonia.seneviratne@ethz.ch	
OMIP	Ready	http://data1.gfdl.noaa.gov/nomads/forms/core/COREv 2.html http://amaterasu.ees.hokudai.ac.jp/~tsujino/JRA55-do- v1.2/ CORE (Ready); JRA55-do v1.2 (Ready)	Yes	Gokhan Danabasoglu gokhan@ucar.edu	
PMIP	Ready	https://pmip4.lsce.ipsl.fr/doku.php; ?	Yes	Masa Kageyama Masa.Kageyama@lsce.ipsl.fr	
RFMIP	Ready	0.4 (2017-01-18)	-	Robert Pincus <u>Robert.Pincus@colorado.edu</u>	
ScenarioMIP	Ready / In Prep.	Land-use – 2.1f (2017-10-05); emissions (in prep.)	-	Detlef van Vuuren <u>Detlef.vanVuuren@pbl.nl</u>	
VolMIP	Ready	3.0 (2017-10-04); EVA module (Ready – GMD below) https://doi.org/10.5194/gmd-9-4049-2016	-/Yes	Davide Zanchettin davidoff@unive.it	

What modeling groups need?

- Description of each experiment
- List of output fields requested from each experiment
- Clearly defined specifications of model output
- Forcing data sets
- Software to help meet CMIP6 data standards and check conformance
 - CMOR3
 - » PrePARE

CMOR3 / PrePARE facilitate (and check) conformance of files to CMIP6 requirements

- CMOR3: (Climate Model Output Rewriter 3)
 - Code for writing model output following CMIP6 specs
 - Code available at https://github.com/PCMDI/cmor
 - Documentation available at http://cmor.llnl.gov/
- PrePARe (Pre-Publication Attribute Reviewer)
 - Code for checking some metadata for compliance with CMIP6 specs
 - Call the CF-checker to check compliance with the CF-conventions
 - See https://cmor.llnl.gov/mydoc_cmip6_validator/

Status

- Major development completed about a year ago
- Several enhancements made this past year
- CMOR being used by CMIP6 modeling centers, input4MIPs, obs4MIPs
- PrePARE being used by modeling groups not using CMOR3 and ESGF to screen out non-conforming datasets

What do scientists need to perform research based on CMIP6 results?

- Output uniformly structured with machine-interpretable metadata
 - Data request, specifications, and CVs
- Easy access to model output: ESGF
 - Data catalog
 - · CoG search
 - replication (Synda)
- Documentation of models & simulations: ES-DOC
- Easy access to errata information: ES-DOC
- Assign PID's and DOI's to datasets to facilitate tracking and to provide credit to modeling groups
- Publication registration service (metric of CMIP6 impact)

What do scientists need to perform research based on CMIP6 results?

- Output uniformly structured with machine-interpretable metadata
- Easy access to model output: ESGF
- Documentation of models & simulations: ES-DOC
- Easy access to errata information: ES-DOC
- Assign PID's and DOI's to datasets to facilitate tracking and to provide credit to modeling groups
- Publication registration service (metric of CMIP6 impact)
- Server-side computation needed
 - subsetting
 - simple reduction (climatology, zonal mean, etc.)

ES-DOC for CMIP6 status



- CMIP6 documentation scope:
 - WGCM has a responsability to document data output for users beyond the usual WGCM science community - this is a key issue for many stakeholders
- Designed so that process is easier for modelling groups:
 - Large fraction is automated
 - Option to start model description from CMIP5 version
 - Modular and agile process
 - Documentation for all steps (+ published WIP white paper)
- Community review :
 - Science contents of model documentation (realms, short tables) on-going (we need WGCM help to identify more science reviewers)
- Beta testing phase on-going (GFDL, IPSL and CCMA, IITM, MPI soon)
 - 20 liaison out of 34 groups identified (we need WGCM help to identify the others)
- · Time line:
 - Science contents of realms ready Nov 1st (V1.0)
 - iPython notebook entry tool with CMIP5 seeding ready Dec 15th
 - Cdf2cim tool ready for ESGF ingestion
 - Community support tools (checklist, training, webcasts, ...) ready Dec 15th
 - Jan 1st for full end-to-end release for model and simulation documentation
 - Challenges due to reduced funding (IS-ENES gap)



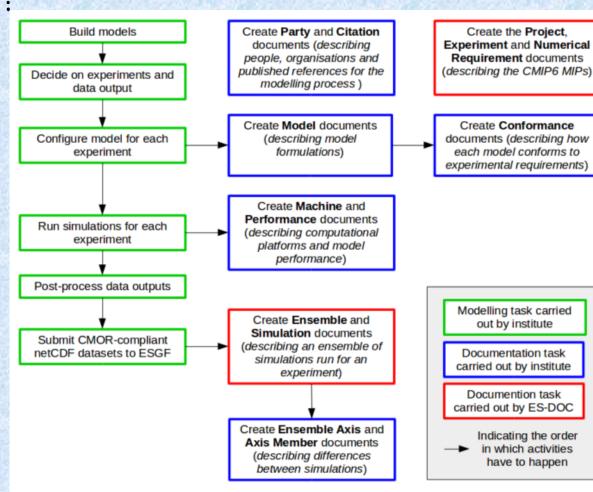


ES-DOC for CMIP6 status



The documentation workflow:

- About half of the documents automated or ES-DOC generated
- The others produced by groups when ready
- Links from "futher_info_URL":
 - Institute's general homepage
 - Description of the experiment
 - Scientific description of model
 - Description of the ensemble
 - Institute's own page
 - Dataset errata information
 - Citation information
 - Performance
 - Datasets in ESGF
- Note: conformance document will capture exact forcing used by groups







What do scientists need to perform research based on CMIP6 results?

- Output uniformly structured with machine-interpretable metadata
- Easy access to model output: ESGF
- Documentation of models & simulations: ES-DOC
- Easy access to errata information: ES-DOC
- Assign PID's and DOI's to datasets to facilitate tracking and to provide credit to modeling groups
- Publication registration service (metric of CMIP6 impact)
- Server-side computation needed

New ES-DOC errata service



- Records issues (problems) with published datasets
- Provides service for responding to queries about datasets identified by their "persistent identifiers" (PIDs)
 - Datasets are labeled with "persistent identifiers" (PIDs)
 - User can determine whether a queried version of dataset/file is safe to use or is
 - affected by an unresolved issue.
 - Has been superseded by a newer version

• In development:

- Exposure of errata service to other services (such as the ESGF CoG front-end and Synda) to ensure real time, automated feedback on data status.
- Incorporation of the issue declaration process in the conventional publishing workflow.





What do scientists need to perform research based on CMIP6 results?

- Output uniformly structured with machine-interpretable metadata
- Easy access to model output: ESGF
- Documentation of models & simulations: ES-DOC
- Easy access to errata information: ES-DOC
- Assign PID's and DOI's to datasets to facilitate tracking and to provide credit to modeling groups
- Publication registration service (metric of CMIP6 impact)
- Server-side computation needed

Citation and data tracking

- DOI's will be assigned at a fairly high level (model/experiment?)
 - Data granularity: DataCite DOI together with citation reference will be assigned to the collection of data from a single experiment and model
 - A reasonably short list of DOI's plus citation reference can be included in publications.
 - Main requirement: ensure proper citation of data acknowledging contributions by modeling groups
- Persistent IDs (PIDs) will be assigned at fine granularity
 - Data granularity: PIDs are assigned to
 - Each CMIP6 NetCDF file during the ESGF data publication and
 - The collection of files comprising an atomic dataset
 - Web service planned for recording lists of PIDs along with citation info. for CMIP6 publications.
 - ES-DOC errata services will be PID-based

ESGF Face-to-Pac Potential use of PIDs in replication workflow.

What do scientists need to perform research based on CMIP6 results?

- Output uniformly structured with machine-interpretable metadata
- Easy access to model output: ESGF
- Documentation of models & simulations: ES-DOC
- Easy access to errata information: ES-DOC
- Assign PID's and DOI's to datasets to facilitate tracking and to provide credit to modeling groups
- Publication registration service (metric of CMIP6 impact)
- Server-side computation needed

Users of CMIP data are obliged to record publications as in CMIP5.

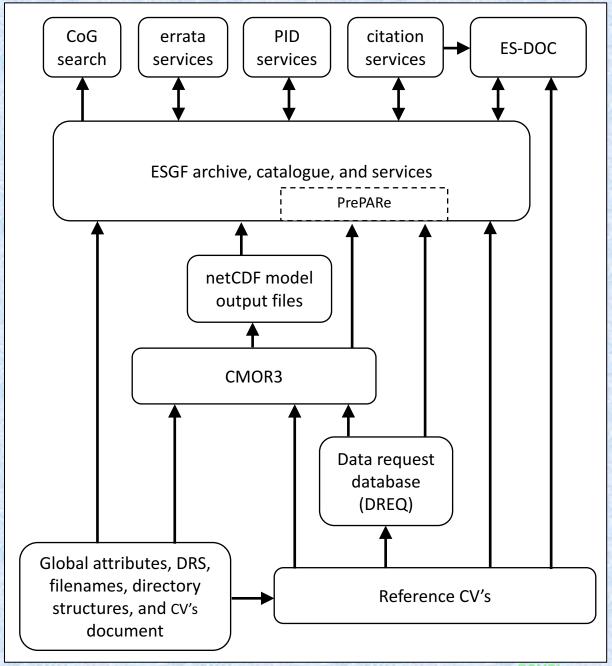
- PCMDI maintains a web-based service for recording publications based on CMIP output
 - https://cmip-publications.llnl.gov/
- This has been recently improved to facilitate entering input
 - Enter DOI and pre-populate most information
- We collect additional (non-mandatory) input identifying the output used, including which experiment, models, and variables
- We plan to provide functionality for recording a list of "tracking i.d's" documenting the data relied on by a study.
 - Permanent record of provenance
 - Can be used to meet government and scientific mandates to make data available.

What do scientists need to perform research based on CMIP6 results?

- Output uniformly structured with machine-interpretable metadata
- Easy access to model output: ESGF
- Documentation of models & simulations: ES-DOC
- Easy access to errata information: ES-DOC
- Assign PID's and DOI's to datasets to facilitate tracking and to provide credit to modeling groups
- Publication registration service (metric of CMIP6 impact)
- Server-side computation needed
 - subsetting
 - reduce volume prior to data transfer (climatology, zonal mean, etc.)

CMIP6 infrastructure has many interdependent components that must effectively work together.

As each component evolves, must be careful not to disrupt other components.



An accurate "dependency" table needs to be constructed, and updates should not be made to a infrastructure component before checking that it remains compatible with its dependent components.

	HAVAN DE MA		A TANK TO STANK WALL		Maria Commission (N. W.		Bally Life and the	Will be the second	Committee of the commit	W.W. Barrell
Infra- structure		Infra-structure element dependent on others								
element needed by others	CMIP6 "specs" doc*	CMIP CVs	DREQ	CMOR & PrePARE	ESGF	ES-DOC	Citation Services	Errata Services	PID services	Long- Term Archival
CMIP6 "specs" doc*	6.2.5 (9/14/17)		6.2.?	6.2.4	6.2.4	6.2.4				
CMIP6 CVs	6.2.4	6.0.0.23 (6/27/ 17)	6.0.?	6.0.?	????	6.0?				
DREQ		6.0.0??	1.00.09 (5/9/ 17)			1.00.??				
CMOR & PrePARE				3.2.3 (4/19/ 17)	3.2.?					2.5?
ESGF			??		2.5? (xx/xx/ xx)	2.5.13		2.5.13		
ES-DOC						0.9.8.0 (xx/xx/ xx)	???			
Citation Services		6.0.0.23					1.0.0 (1/18/ 17)			
Errata Services								0.4.0.0 (xx/xx/ xx)		
PID services								1.0.0 (latest)	1.0.0 (7/11/ 17)	
Long- Term Archival						у.у.у	1.0.0	у.у.у		y.y.y (xx/xx/ xx)

5 December 2017

CMIP6 infrastructure

- "Crunch time" is here.
- A cleanly operating simple infrastructure is better than a more powerful one that has bugs
- The climate science community is grateful for all the hard work going into ESGF.